DOCUMENT RESUME

ED 134 539 95 SP 010 702

AUTHOR	Pottinger, Paul S.; Klemp, George O.
TITLE	Concepts and Issues Related to the Identification, Measurement and Validation of Competence.
INSTITUTION	McBer and Co., Boston Mass.
REPORT NO	R
PUB DATE	76
NOTE	83p.; Prepared by Institute for Competence Assessment of McBer and Co., Boston
AVAILABLE FROM	McBer and Company, 137 Newbury Street, Boston, Massachusetts 02116 ($3.50)
EDRS PRICE	MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS	Academic Achievement; Cognitive Objectives; *Educational Assessment; *Evaluation Criteria; Evaluation Methods; Evaluation Needs; General Education; Liberal Arts; Measurement Instruments; *Measurement Techniques; *Performance Based Education; *Post Secondary Education; *Relevance (Education); Test Validity

ABSTRACT

This report reflects experiences in the development of new conceptual frameworks for defining learning outcomes that are most desirable for effective life preparation and that emphasize the way people process and integrate information rather than how well people merely store and retrieve information. These new measures, for use in institutions of postsecondary education, must: (1) be sensitive and relevant to important learning outcomes; (2) have general significance to a wide variety of career and life outcomes; (3) have practical utility for educators; (4) be methodologically and technically innovative, e.g., utilize operant rather than respondent behaviors; and (5) be quantifiable and thus amenable to rigorous determination of reliability, validity, and meaning. Keeping in mind these concerns, innovative measures have been developed that attempt to answer the need for more operant measurement techniques to assess the factors of process, integration, and implementation. These measures are described and organized according to their particular outcome domains--cognitive, effective, or social. Of these tests and measures described, none is especially useful as a diagnostic or assessment tool in isolation. Thus, the General Integrative Model, involving the use of several of the tests and measures described, is introduced as one way of evaluating the meaningful integration of life skills. (MM)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED134539

INSTITUTE FOR COMPETENCE ASSESSMENT

ICA

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Paul A. Pottinger

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

SP010 702

CONCEPTS AND ISSUES
RELATED TO THE IDENTIFICATION,
MEASUREMENT AND VALIDATION
OF COMPETENCE

Paul S. Pottinger, Ph.D.
and
George O. Klemp, Ph.D.

Institute for Competence Assessment
McBer and Company
137 Newbury Street
Boston, Massachusetts 02116

© McBer and Company
NOT FOR QUOTATION WITHOUT
PERMISSION OF THE AUTHORS

TABLE OF CONTENTS

SECTION I

INTRODUCTION AND GENERAL PROBLEM STATEMENT 1

Summary 7

SECTION II

CONCEPTUAL ELABORATIONS TO CLARIFY PROBLEMS AND SUGGEST SOLUTIONS 9

Critical Concepts in Defining Generic Abilities 10

1. Measuring Use Of Knowledge Rather Than Storage of Knowledge 10
2. The Problem of Method Variance 12
3. Generic Skills Vs. Observable Performance Skills 14

Empirical Linkages Between Educational and Postacademic Life Requirements 15

1. Criterion Sampling Measures 16
2. Criterion Correlated Measures 17
3. Causally-Related Criterion Measures 20

Determining the Meaning of Measures 22

The Problem of Establishing Criterion Levels of Performance Standards 27

1. The Problem of Maximum Levels 29
2. The Problem of Interactions 32

Implications of New Measures for Policy Research and Decisionmaking 34

1. Improvement of Means/Ends Linkages Could Be Facilitated 34
2. Improved Cost/Benefit Calculations Concerning Any Aspect or Part of the University Become Possible 35

Technologies for Identifying Skills, Abilities and Other Characteristics Related to Competence 38

1. Job Element Analysis 39
2. Behavioral Events Analysis 46
3. The Organizational Climate 48

SECTION III

PROTOTYPE MEASURES OF LEARNING OUTCOMES RELATED TO LIBERAL ARTS AND THE PROFESSIONS 53

Discussion of Measures 54

Measures of Cognitive Outcomes 54

1. Critical Thinking 54
2. Learning Styles 56
3. Programmed Cases 58
4. Analysis of Argument 59
5. Concept Formation 59
6. Speed of Learning 59
7. Savings Score 59
8. Proactive Case Response 60

Measures of Effective Outcomes 60

1. Diagnostic Listening 60
2. Achievement Motivation 62
3. Self-Definition/Cognitive Initiative 63
4. Socialized Power 64
5. Stage IV Power 64

Measures of Social Outcomes 64

1. Nonverbal Sensitivity 64
2. Moral Reasoning 65
3. Affiliation Motivation 67
4. Social-Emotional Maturity 67

A General Integrative Model 68

Characteristics and Advantages of Competency-Based Measures 72

References 76

SECTION I:
INTRODUCTION AND GENERAL PROBLEM STATEMENT

As a part of a broader mandate, the National Institute of Education has solicited proposals to strengthen the scientific and technological foundations of education. These are important concerns for the traditional college student. They are of critical importance to the nontraditional learner who is entering or returning to the educational mainstream in increasing numbers.

The assumption that college attendance prepares one adequately for adult life roles has been called into question in recent years. People are realizing that a college education does not necessarily lead to a greater degree of success in adult life. The once sacred notion that education is a good end in itself is being replaced by the notion that educational institutions must demonstrate their impact on clearly stated learning goals. Students are demanding preparation and credentials that have more meaning in the world of work. Educators are asking for better information to determine what will satisfy these needs.

The issues of assessment and measurement have come to the forefront of education. With regard to students who seek higher education with the hope of fulfilling their expectations for success in work and other life roles, traditional measures of academic success are often of little relevance. Course grades, credit for time in class, and standard aptitude and achievement test scores repeatedly have been shown to be unrelated to demonstrated competence in the postacademic world. The attainment of a degree is now recognized as a measure of "doing time" in the educational process rather than as a measure of achieving clearly specified life-relevant learning outcomes.

Problems in determining criteria for granting degrees and in linking these criteria to adult life roles have created special needs. One need is for new conceptual frameworks to define these problems more clearly. Another need is for more sensitive, valid and relevant measurement techniques. And there is a need for more systematic collection of data in order to answer critical questions of test validity, meaning and relevance.

More than ever, liberal arts educators want to know and need to demonstrate if they are accomplishing the goal of preparing people effectively for adult life roles. The development and use of assessment and evaluation techniques, however, have not kept pace with the need for better answers to these fundamental questions. Changes in the art and science of assessment have lagged behind changes in practice. Higher education needs to make changes in practice. To do this effectively, it also needs to know what changes are warranted; what outcomes are most desirable for effective life preparation; and how progress toward these outcomes can be measured.

Educators have attempted to respond to this challenge with new assessment techniques. Unfortunately, most new measures and methods of assessment, which have broken away from a narrow knowledge orientation, are insensitive to important learning changes; lack reliability, validity and theoretical/empirical bases; and lack relevance to newly articulated goals. Often, they are poorly linked to adult life requirements, are too costly and are methodologically limited. For example, many innovative approaches to assessment are being developed, which borrow from techniques and procedures developed by industrial psychologists, such as:

- portfolios
- journals
- juries
- committees
- life histories
- self-assessments
- supervisor, peer and/or client ratings
- in-basket tests
- work sample tests
- games
- simulations
- rehearsed performances.

Ironically, most of these efforts to break away from traditional measures suffer from many of the same shortcomings of traditional tests. That is, (1) the techniques tend to be highly subjective and open to broad interpretations; (2) they do not easily lend themselves to standardization across institutions or even among individuals who use them; (3) there is as yet little or no empirical evidence that the performances being measured are any more related to success outside of academia than performances measured by traditional measures.

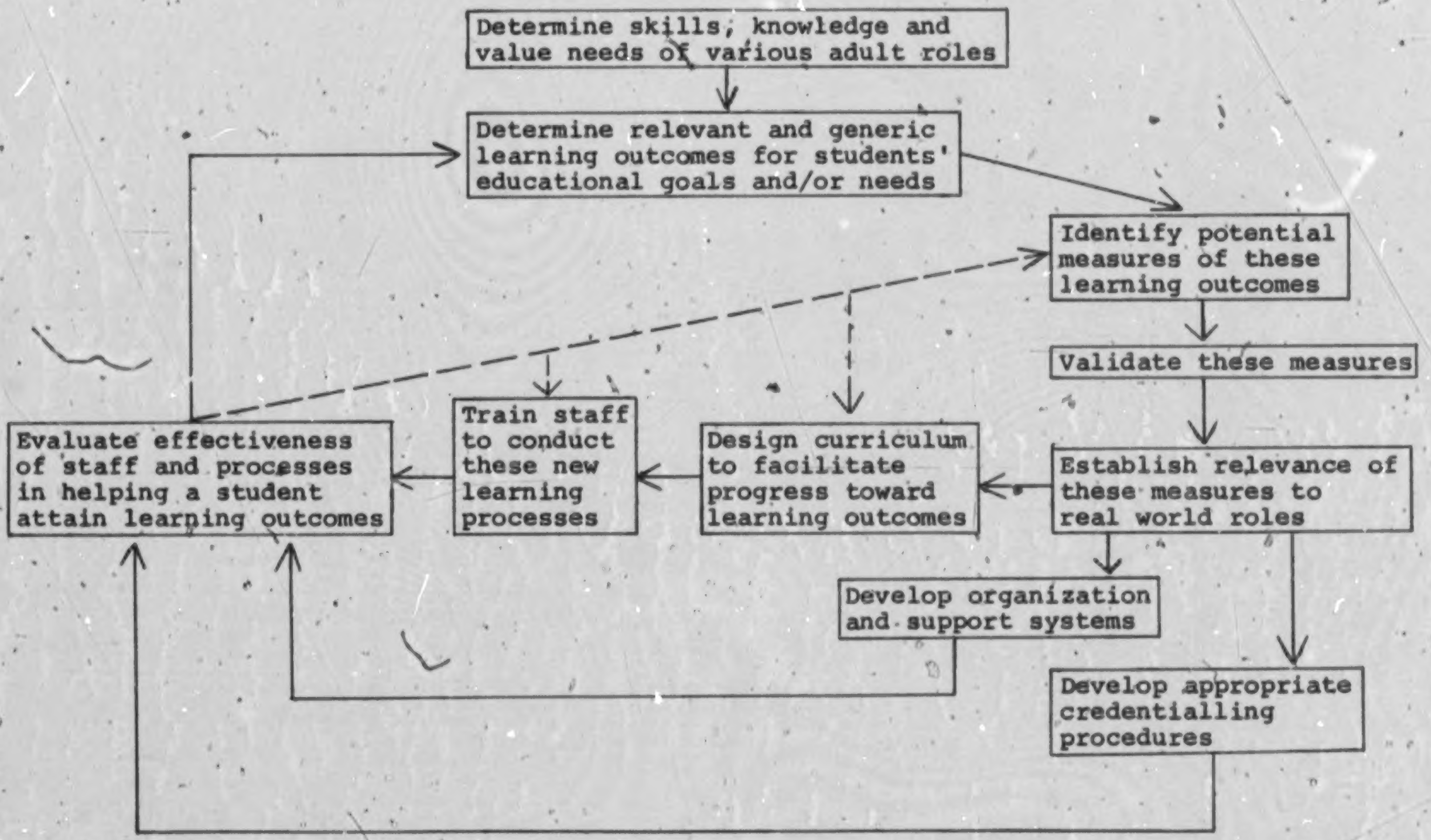
The predominance of new techniques and procedures do not appear to lend themselves to rigorous empirical analysis nor to construct-validation. Rather they only change the focus of subjective judgments about student learning outcomes. Thus, while these innovations have seemed appealing from the point of view of changing values and ideologies, they have lost the rigor necessary for understanding what is really being assessed and how this relates to a student's preparation for life. Reliable and valid gains in knowledge have been forfeited in the processes of broadening techniques and eliminating the irrelevance of traditional assessment methods.

Assessment procedures are always part of a complex synergistic educational system. The development, validation and implementation of new assessment techniques cannot take place in isolation from teaching, curriculum and institutional support systems. One model for conceptualizing the process of implementing changes in assessment procedures appears in Figure 1. This model also demonstrates the central role of assessment in the educational system.

New measures of learning outcomes which are true to the real goals of postsecondary education and sensitive to student progress are needed to enable teachers to calibrate their techniques, to make effective changes in curriculum, and to indicate where there are needed changes in organization and support systems. Such new measures are also needed to convincingly demonstrate the effectiveness of innovative programs. Students, faculty, admini-

Figure 1

Implementation of Changes in Assessment


Determine skills, knowledge and value needs of various adult roles
Determine relevant and generic learning outcomes for students' educational goals and/or needs
Identify potential measures of these learning outcomes
Validate these measures
Establish relevance of these measures to real world roles
Design curriculum to facilitate progress toward learning outcomes
Train staff to conduct these new learning processes
Evaluate effectiveness of staff and processes in helping a student attain learning outcomes
Develop organization and support systems
Develop appropriate credentialling procedures

ICA

strators, higher education supporters, the public and the Congress all need to know that innovations are working effectively. Policy makers at every level are eager to know what works and why. Because standard methods of educational evaluation measure a limited and specialized type of learning outcome that turns out not to be related to important life requirements for occupational success or life adjustment, these standard evaluation results have been poorly utilized by curriculum developers, program evaluators, or policy makers at any level. While educational innovations may importantly affect learning outcomes, these outcomes simply cannot be measured in traditional ways or with traditional tests (see McClelland, 1973 for a discussion of the evidence). Yet, one of the major difficulties in trying to revitalize postsecondary education is that any changes made tend to be evaluated in terms of traditional academic tests.

The deficiencies of assessment methods in higher education are not due to lack of talent, commitment, or dedication among educators. Nonetheless, ideas which seem good in the abstract are often too difficult for practitioners to make functionally useful. Thus, faculty tend to fall back on traditional measures or subjective judgments by default. Some educators do not know what questions to ask or how to ask them in ways that can lead to productive results. Many educators also do not understand technological and methodological issues involved in clarifying goals and measuring progress toward them. The importance of measuring outcomes of generic cognitive and noncognitive skills is often overlooked or poorly understood in higher education in spite of the concern of postsecondary institutions for the development of general abilities. (For an elaboration of critical concepts in assessment see Section I.)

Summary

1. New conceptual frameworks are needed for defining learning outcomes that are most desirable for effective life preparation. These conceptual models must emphasize the way people process and integrate information and implement solutions to problems rather than how well people merely store and retrieve information.

2. Better techniques for developing measures which tap relevant learning outcomes are needed. They must emphasize the quantification of outcome criteria so that educators can rigorously and meaningfully validate these measures. They must emphasize the new methods of assessing learning behaviors apart from the predominantly passive or respondent methods now in use.

3. Practical methods for validating new measures are necessary so that institutions understand the meaning of their assessment measures and techniques. These methods should include construct-validation.

4. These measures must be referenced to criteria which reflect requirements for success in the postacademic world, if the real meaning of new measures and techniques are to be relevant to the assessment of one's preparation for work and other adult life roles.

These relationships must not be mere correlations between observable behavior and successful outcomes, but they must reflect causal links between learning and successful outcomes.

5. Measures are needed which (a) are sensitive to student changes, (b) provide useful feedback about the progress they are making toward their own learning goals and, (c) enable teachers to develop and evaluate better curriculum and teaching techniques.

6. Program effects on learning must be compellingly demonstrated. Construct-validated and criterion-referenced measures must be utilized to show that innovative practices of postsecondary education are effective.

SECTION II:
CONCEPTUAL ELABORATIONS TO CLARIFY PROBLEMS AND SUGGEST SOLUTIONS

ICA has had considerable experience in identifying, defining, measuring and validating generic cognitive abilities and non-cognitive skills. ICA's development of assessment techniques in institutions of postsecondary education and in professional occupational institutions and organizations has been fairly unique.

The discussion in this section will reflect these experiences as well as the need for a fuller perspective on critical concepts, practices and assessment techniques.

These conceptual elaborations will cover the following six areas:

- Critical Concepts in Defining Generic Abilities;
- Empirical Linkages Between Educational Assessment and Postacademic Life Requirements;
- Determining the Meaning of Measures;
- The Problem of Establishing Criterion Levels or Performance Standards;
- Implications of New Measures for Policy Research and Decisionmaking;
- Technologies for Identifying Skills, Abilities and Other Characteristics Related to Competence.

Critical Concepts in Defining Generic Abilities

1. Measuring Use of Knowledge Rather Than Storage of Knowledge

Psychologists have often failed to develop measuring instruments that are sensitive enough to detect effects of primary interest to educators. According to McClelland (1976) there is ample reason to believe that educational psychologists have unnecessarily restricted the range of methods they have employed to measure the impact of higher education. Time-saving and money-saving incentives have resulted in a predominance of measures which utilize the multiple-choice questionnaire format or which remain highly subjective and unamenable to determining validity and meaning.

One reason for this is that traditionally educators have limited their focus in teaching (and assessment) on the transmission of knowledge (i.e., course content). The rhetoric of higher education regarding liberal arts education has reflected the objectives of students becoming critical and discerning thinkers, competent problem-solvers, and socially mature and responsible citizens. Yet predominantly, assessment techniques have been limited to determining students' abilities of retention and recall of subject matter.

It would serve us well to ask the extent to which our current assessment techniques have any bearing on what people do in real life and on the competencies that they possess. In our daily lives we are constantly called upon to process various kinds of information, to analyze its components, to associate this new information with that which we have stored away in our memory, to partial out the crucial information from the trivial and to integrate this information into our cognitive structure. In this way, we constantly use information from many sources to solve problems, and in the process we learn new things about our world and ourselves. In truth, people are almost never asked to recognize a correct answer among a list of three or four alternatives. Rather than being reactive to such well-defined situations, people must be proactive in situations which provide only partial information.

The one thing most traditional testing methods have in common, regardless of what they purport to assess, is this: they only measure one's ability to retrieve information after it has been stored. Many such methods fail even in this. A multiple-choice test, for example, measures the ability to recognize rather than recall. Essay tests are very subjectively scored, even when there is only one *correct* answer or line of reasoning as is often the case.

Storage and retrieval of information are not the important issues for higher education. Indeed, Ebbinghaus demonstrated many years ago that 70 percent of that which is learned in the classroom is forgotten within one year. Rather, the issue is a more substantive one: how is the knowledge gained in coursework used to come to grips with the practical problems of living? Implicit in this are three related issues of particular importance: how able are people in processing new information for problem solving; how able are they in integrating this information to form new solutions; and how able are they in implementing these solutions? Little wonder that test scores, grades and credentials based on retention and recall of facts correlate so poorly with demonstrated competence in the world of work and adult life in general.

While cognitive processing and integrating skills and important noncognitive skills are often learned in higher education, teaching and curriculum often do not relate directly to these abilities in a clearly articulated fashion; nor do assessment procedures tap these abilities in any rigorous quantifiable fashion.

2. The Problem of Method Variance

Intuitively, the reason tests have been avoided for so long is that it has been known that only a small part of the richness of thinking and behavior is tapped by paper and pencil tests.

There are many qualities that educators would like to measure, such as common sense, managerial skills, leadership behavior, interpersonal effectiveness, moral reasoning, and initiative. Unfortunately, educators have to settle for measuring small components of these qualities in terms of specific knowledge, skills and abilities that they hope are related to these more general qualities. One reason for this reduction in measurement is that the technology of ability measurement is not good enough to get at the larger more consequential characteristics of people.

We can easily fall prey to further reductions in the quality of assessment by limiting ourselves to only one method of measurment. Campbell and Fisk (1959) have documented the common sense notion that the more one increases different perspectives and techniques in measuring a phenomenon, the better will be the measurement. Traditionally, in measuring learning phenomena, we have limited ourselves to a set of respondent-type measures. These measures typically require multiple forced choices among a set of prepared alternatives in a paper and pencil format. By limiting ourselves to these paper and pencil tests, we are measuring the effect of the test format as much as we are measuring the knowledge, skills and abilities being assessed. In technical terms, this is the issue of "method variance," i.e., how much we are measuring the method relative to how much we are measuring some personal attribute.

Assessing different areas of academic ability by using a series of paper and pencil tests is analogous, for example, to measuring how fast someone can drink by requiring one to use a straw. In this example the paper and pencil test and the straw are equivalent in that they both limit the phenomenon being measured in a reliable way. We would get a better understanding of true academic ability, as well as the ability to drink quickly, if we worked toward eliminating the constraints of measurement. One way of doing this is to utilize measures that break away from single modes of measurement. In doing so, in any case, we must require that the measurement techniques we use are objective and quantifiable.

We will discuss in Section III a number of measures which differ in their perspectives. These measures move toward the elimination of method variance as a confounding factor in measurement while remaining objective and quantifiable.

3. Generic Skills vs. Observable Performance Skills

A third concept has to do with measuring abilities that are causally related to successful performance rather than being merely correlated with perfofmance. This point will be elaborated in the next part of this section. Suffice it to say here that many assessment techniques are based on external behaviors which, although they are the building blocks of successful performances, tend to be reductionistic and lack meaning because they fail to assess the underlying causes of these behaviors. This often results in the assessment of a laundry list of behaviors which may have little generalizability in or transferability to a variety of real life requirements. This problem has important implications for teaching and curriculum as well as for problems of assessment because often observable but superficial behaviors rather than these causal underlying factors are taught. Thus, what is actually learned, as well as what is assessed may have little general significance in post-academic life.

Empirical Linkages Between Educational and Postacademic Life Requirements

Let us look more closely at this problem of causally related measures as we elaborate on concepts germane to linking educational assessment techniques with postacademic requirements for success.

At the heart of the issue of linking assessment to the postacademic world is the notion of criterion referencing. Many of the measures which fail to predict performance outside of academia, e.g., intelligence, scholastic aptitude, verbal proficiency, and the like do so because they are norm-referenced. The distinction was well defined recently by Messick (1975):

> A norm-referenced test is one that is constructed to yield test scores that discriminate among individuals on the trait measured by the test and that are interpretable in terms of the relative performance of other individuals and groups on the same test. A criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards. (underscoring mine)
>
> At the level of interpretation, the distinction seems clear: A norm-referenced interpretation compares an individual's test performance with the performance of others, whereas a criterion-referenced interpretation compares it with a performance standard.

It is easiest, perhaps, to understand the importance of criterion referencing for linking educational assessment techniques with the postacademic world if we examine the use of assessment measures in the world of work.

For convenience of discussion and analysis we will arbitrarily categorize techniques into three basic types of measures and procedures which fall somewhere along a continuum of most

ICA

to least *directly performance related*. At one end of the continuum are criterion sampling measures which consist of transfering on-the-job behaviors directly into the assessment situation. At the other end are measures which can be demonstrated to be statistically related to work performance, although the reason for this relationship (correlation) is not clear. Somewhere between these two extremes are measures causally related to performance criteria, although they do not involve direct criterion sampling. All these tests are in some sense criterion-referenced, but this fact alone is no guarantee that the test will be highly predictive of performance criteria or will allow one to draw appropriate conclusions about educational strategies.

We will examine assessment techniques as they relate to management and leadership roles since these, perhaps, reflect the major general learning outcomes espoused by liberal arts.

1. Criterion Sampling Measures

With regard to complex managerial and leadership roles, the assessment center approach is a popular example of this type of measure. One of the major attractions of the assessment center approach is that it is more job performance related than ordinary test batteries, performance records, etc.; that is, it samples behaviors required in management, or at least analogous to the work itself, through such techniques as management games, leaderless groups and simulated work samples (e.g., in-basket exercises). The attempt to predict complex leadership and management behavior through procedures that are directly performance related--the essence of the assessment center approach--is, of course, the major strength of this technique. However, while such direct assessment procedures are observably performance related, they lack validity because the behavioral observations suffer from all the vagaries of subjective-rater biases, and the behaviors observed are often unreliable (or rarely examined for reliability). Both performer and rater reliabilities, then, tend to be low (if measured at all) and therefore greatly diminish the validity of these techniques. Furthermore, direct performance observation and assessment techniques are time-consuming, labor-intensive, costly, and less amenable than other techniques to quantification and statistical treatment. In general, behavioral sampling techniques can be of great value if care is taken to assure their objectivity and reliability.

2. Criterion Correlated Measures

The instruments in this category include paper and pencil tests which measure psychological constructs. From the test scores of those being assessed, assumptions or predictions are made about how one might perform in a variety of situations. These tests range from those that try to predict specific behaviors in limited situations to broad trait measures which supposedly tap some enduring attributes of personality or character that prevail in at least all normal situations. A proliferation of examples could be used here, since assessment technology from its earliest years has focused most heavily on correlational techniques. Intelligence tests, personality tests, and standard attitude and achievement tests are the most common examples of this type of measure based on the technique of empirically retaining a subset of a massive group of items such that the items that remain differentiate between criterion groups. But in general the correlation between tests of this type and performance criteria, though statistically significant, account for very little predictive power (e.g., a typical correlation coefficient of .30 translates to only nine percent of real predictability).

Indirect measures, including those just mentioned, often have high performer and rater reliabilities. They also tend to be efficient, objective, inexpensive and highly amenable to statistical analysis and treatment. However, they often lack validity because they are vague or unrelated to (unpredictive of) actual performance. For example, Ghiselli (1965) conducted an exhaustive review of predictive studies for an impressively wide variety of jobs and occupations in the U.S. using an equally impressive array of tests and measures. Taking all jobs as a whole, the average of the maximal performance predictive validity coefficients was a meagre .33. Conversely, taking all tests' categories as a whole, the highest grand average performance predictive validity coefficient was .30. Obviously, matching the *right* test battery with the *right* job enhances these averages, but not impressively. Furthermore, while some tests reveal significant construct validity coefficients, our interest is primarily in predictive validity where the relationship measured is between test scores and performance (not just test scores and other test scores).

Before addressing ourselves to the third category of measures and procedures, a caveat comparing the top and bottom of the continuum in terms of effectiveness in predicting quality performance is in order. While the assessment center approach has been an appealing possibility for alleviating many of the problems of management and leadership performance prediction in spite of its costly and time-consuming characteristics, this approach has not yet consistently been demonstrated to be more effective than paper and pencil tests combined with subjective supervisor assessments of past experience and performance, experience records and the like (Wilson and Tatge, 1973). For example, these authors summarize data comparing assessment center ratings with paper and pencil tests of intelligence, ability and personality. They report that, at best, assessment center ratings increase predictivity of standard personality measures by too small an increment to justify the cost. In fact, the authors report that this costly combination of procedures does not predict as well as scores based on a battery of tests and background information.[1]

---

[1] The evidence for this conclusion by Wilson and Tatge was a comparison between a "best ease" assessment center study done by Wollowick and McNamara (1969) and the predictive study of management performance at Standard Oil of New Jersey reported

Wilson and Tatge's explanation for this lack of improvement in prediction by direct performance observation and ratings comes from extensive research which shows that the critical measures in assessment centers relate primarily to a candidate's skill and sensitivity in interpersonal relations. Such characteristics as forcefulness, dominance, passivity, dependence, nonconformity, orientation to work, self-confidence, energy level, persuasiveness, need for approval, etc. are also commonly measured by paper and pencil tests, patterned interviews and systematic interpretations of records of past experience.

Thus, while we must preserve the essence of the assessment center approach to obtaining validated performance-related measures, we must also capitalize on the objectivity, reliability, and efficiency of more standard types of measurement techniques while maximizing predictive validity.

3. Causally-Related Criterion Measures

Another variety of assessment techniques and procedures exists which draws from the strengths of the other two categories while minimizing their weaknesses. These tests or procedures, in other words, are clearly related to performance while simultaneously being objective, reliable and efficient and amenable to statistical analysis. This category is often referred to as competency-based measures and procedures.

---

in Tagiuri (1961) using tests and background information. The authors concluded that even when scores from assessment center ratings are combined statistically (rather than clinically), they still fail to exceed similar combinations of tests and personal history data.

A major assumption of this approach is that knowledge, skills and abilities that can be defined objectively are seldom sufficient indicators of how well a person will perform on a job, either at the entry level or in the future. There are many other factors that relate to performance but are not tapped by traditional assessment techniques, such as motivation, observation abilities, empathy, tenacity, the ability to think clearly under stress, the ability to anticipate, analyze and solve problems, and many others. Often these factors are intuitively obvious as critical to managerial and leadership success, but rarely measured effectively if at all. It is these and other variables related to complex higher order management and leadership abilities that causally-related criterion measures are designed to assess.

The focus here has been on the development of measures which will predict competent performance in managerial and leadership roles. This discussion reflects the work of ICA in the world of work, but it should be apparent that the types of skills, abilities, and other characteristics required of effective performance in these roles are similar to or consistent with the goals of higher education in preparing students for the world of work and for life in general.

Everyone *manages* and *leads* something or someone--if only oneself--in adult life. Clearly, educators as well as employers need a better understanding of what constitutes sound management (e.g., critical thinking, problem solving) and effective leadership (e.g., the ability to implement effective solutions). Furthermore, better ways are needed to teach and assess the causal factors that underlie these characteristics of adult roles in life. The concept of causally-related measurement is as critical to education as it is to the world of work, and it provides a framework for making better empirical links between education and the postacademic world with respect to teaching, curriculum and relevant learning outcome assessment.

Making more direct links between education and work is important because students want better preparation for occupational roles; but it is equally, if not more, important because the goals and outcomes of liberal arts education need to be empirically demonstrated as congruent with and causally related to success in work and life in general.

Determining the Meaning of Measures

As background to this discussion, we have already stressed the importance of changing the focus of assessment from merely asking for recall and recognition of content to measuring how one processes and utilizes this information. If assessment techniques are to have sufficient meaning and credibility for determining if students are adequately prepared for life, we can no longer be satisfied with content-valid tests. Construct validation must be determined. Furthermore, we have stressed the importance of creating criterion-referenced measures which are predictive or reflective of real world requirements for success. The following discussion (Pottinger and Klemp, 1976) is a further elaboration on the necessity for construct validation and empirical linkage of measures to obtain maximum meaning of what is being assessed.

Messick (1975) has argued that, until measures have been construct validated, they lack the meaning essential to utilizing them as instruments of general educational theory. McClelland (1973) further argues that, until construct validated measures use relevant real world events among their criterion referents, their value in assessing preparedness for work and life is limited. Educators have often failed to pay attention to construct validity because they "view desired behaviors as ends in themselves with little concern for the processes that produce them or for the causes of the undesired behaviors to be rectified" (Messick, p. 959). In other words, "construct validity is not usually sought for educational tests, because they are typically already considered to be valid on other grounds, namely, on the grounds of content validity" (Messick, p. 959).

In short, educators have traditionally been satisfied with knowing that the content of tests adequately sample a class of situations or subject matter. Messick (1975) argues that content validity does not provide an evidential basis for interpreting the meaning of test scores, and McClelland (1973) argues further that the interpreted meaning of scores that come from construct validation must be strengthened by tying these constructs directly to the world of events outside of academia.

The theoretical distinction between general education and competency-based education is that the latter requires an empirical and causal link between measurement responses and their meaning, as related to real world life outcomes. Most competency-based programs, however, merely correlate test responses with specific criterion-referenced outcomes (and many do not even do this) without discovering the underlying causes of these responses. Many educators make the mistake of thinking that if a test correlates with a behavioral criterion variable in the world of work or elsewhere outside of the academic world, one can develop competence by "teaching to the test." But this notion confuses correlation with causation, i.e., the fact that tests correlate with observable criteria may only indicate the existence of a causal intervening variable which is really responsible for behavior and which has not been measured.[2]

Clearly the mandate for competency-based postsecondary education is to identify skills and abilities that produce

---

[2] For example, vocabulary is correlated with college grades. However, one would not go about improving college grades merely by increasing vocabulary. Doing well in school requires abilities for problem solving, utilizing new information, and other skills not measured by vocabulary tests. Vocabulary is merely a tool, and how it is used depends upon other abilities and characteristics of the individual. One cannot do well in school without a reasonably adequate vocabulary, but having a strong vocabulary will not guarantee success in school without its effective use.

(cause) desired outcomes; to develop curricula aimed at the acquisition of these skills and abilities; and to design and validate measures that are sensitive to the acquisition processes and are representative of the criterion outcomes. One should not consider curriculum development apart from assessment issues and neither should be considered in the absence of identified valid performance criteria. Only when these conditions are satisfied does it make sense to "teach to the test."

The skills tapped by genuine competency-based tests (i.e., causally-related criterion measures) are largely independent of the content areas in which they are used. For example, the tests for thematic analysis, analysis of argument, problem solving, speed of learning, and other such measures described in the next section test for generic abilities (competencies) which can be demonstrated in the context of any specific content area. These tests can be adapted to the natural sciences, social sciences, and humanities with equal facilitiy; the content area does not determine the effectiveness of the test. We will always need tests of knowledge, but we also need tests of the way this knowledge is used. The measures discussed in the following section satisfy both of these criteria, which represent the essence of competency-based assessment.

Common criticism leveled at the competency-based education movement is that its focus is by definition limited to preparation for specific vocations. A narrow correlational model of competence has fostered this notion, and this concern is legitimate to the extent that criterion validities depend exclusively upon specific job-oriented criterion reference groups. Such validities for liberal arts or general education "are of sporadic interpretive utility" at best since they ignore the linking of test behavior to a more general attribute, process, or trait which provides an evidential basis for interpreting the *processes underlying test scores*. (Messick, 1975)

We strongly endorse this position, but hasten to add that construct validation is itself all too often limited in the types of referents it uses to provide meaning to test scores. Thus, we advocate a validation model that draws from the strengths of construct validation more heavily in the context of *real world events or life outcomes* than in the context of other constructs alone or "laboratory" behaviors. While Messick (1975) de-emphasizes criterion-referencing, he only does so (1) in terms of using criterion-referents outside of the context of construct validation and (2) perhaps in terms of the type of criterion used as referents. Indeed, all validation is criterion-referenced. The difference in criteria (e.g., "real world" performance, other tests, or observable "laboratory" behavior) determines the extent to which the meaning of the test responses are general or specific and of theoretical or real world significance. A difference between McClelland's (1973) and Messick's point of view is McClelland's emphasis on choosing real world behaviors as opposed to tests (which typically tap respondent rather than operant behaviors) and laboratory behaviors, as criterion referents. Thus, criterion-referents constituted by a nomological network of life outcomes are consistent with Messick's argument. Espousing such referents differs from Messick's point of view only in terms of emphasizing their selection as criteria for construct validation, not in the validation procedures or concepts themselves. In other words, Messick's notion of construct validation theoretically would include criterion behaviors, but empirically there are differences in emphasis on the types of behaviors to be included. It is for the sake of this difference in emphasis, not theoretical differences, that we have isolated real world events or life outcomes as critical factors in determining the real meaning of tests.

The strength and future of competency-based education rests on its ability to support the rigorous type of research analysis which involves construct validation based heavily upon real world life outcomes. Until we have identified the critical intervening variables in the causal chain between the educational experience and performance outside of academia, we will be legitimately faulted by critics who view competency-based assessment (and education) as too narrow in scope.

The Problem of Establishing Criterion-Levels of Performance Standards

As the meaning of measures becomes established by construct-validation and empirical (criterion-referenced) links between education and the requirements of postacademic life, the question of what criterion levels of performance is necessary for granting credentials is made easier, perhaps, because concrete information exists with which educators can make sound judgments. Yet, the problem of establishing standards for levels of performance is a complex one because (1) this determination of appropriate levels of performance is dependent upon educators' goals for credentialing students, and (2) technical issues related to understanding the meaning of maximum levels of performance and the meaning of complex interaction of abilities probably necessitate highly subjective determinations of criterion standards.

With regard to the first point about determining standards of performance, Hodgkinson (1975) stresses the importance of asking good questions about the use and purposes of assessment. Sound judgment and planning are necessary to avoid proceeding with evaluative decisions based on ambiguous criteria, standards and/or levels of outcomes. These questions must include: *Who establishes criteria or standards*--an external auditing agency, a faculty member, the institution? *What is the reference group with which one will be compared*--performers in the real world, students in past years, other students currently being evaluated, one's own past performance, an *ideal* student? *What is the proper method of comparison*?--norm-referenced tests, criterion-referenced tests, behavioral measures, narratives (e.g., portfolios, diaries of past experience), unobtrusive measures, etc.? What is the nature of the standard--job performance in the "real world," individual growth and development, ideological ideals of performance, standardized scores? What is the function of the standard -- to select or reject people, to improve performances, to admit students to professional schools or jobs?

If these questions are asked and the answers are concrete, specific and meaningful, a student should know who is judging him, how he will be judged, the nature of these judgments, the objectives related to them, and how well he must perform to meet those objectives.

With regard to the second point about determining standards of performance, two conceptual or technical considerations are also relevant.

1. The Problem of Maximum Levels

Credentials are often restricted to those whose scholastic performance and/or test scores are higher than minimal levels required for work or other social roles. Such occurrences discriminate unfairly against those who are competent to work, for example, but who are selected out of occupational opportunities by those who believe in the simple equation: higher academic achievement means better work or life performance. The tacit assumption that superior abilities in all measured characteristics are necessary or even desirable for performance is highly questionable.[3]

---

[3] A simple motor skill example will demonstrate this point:

Measures typically used to assess job task performance and performance relating to the mastery of units in a curriculum typically have little bearing on how subunits interact. For any given job, life task, or individual performance, component skills in one area can compensate for deficiencies in others creating a variety of combinations of individual performance levels which could theoretically add up to equivalent overall performance. Thus, minimal levels of performance on individual variables (which compromise overall competence) may have little meaning by themselves. Their interactions with respect to outcomes may have far greater significance.

We are most familiar with this problem in cognitive areas of education. We are often taught language use, verbal reasoning, spatial relationship, reading comprehension, abstract reasoning and syllogistic analysis (e.g., as measured by Miller Analogies) as discrete units of curricula. Assessment of integrated or general skills such as problem solving often do not take into account the interactive nature of skills in these subcomponent areas. Cognitive measures are used almost exclusively in assessment as if the qualities they measure did not interact, i.e., they are tested separately.

The importance of interactions, while intuitively obvious.

---

we know that an automobile driver must grip the steering wheel with enough force to maintain control of the car. But beyond a certain level of pressure, added strength in holding the wheel does not increase overall driving competency. And this is just one of some 3,400 discrete behaviors identified by researchers as making up the task of "driving."

in the motor skills area, have not been carefully attended to in cognitive and social/emotional areas of assessment. Yet, once individuals have gone through a series of academic life experiences that enhance their competence in dealing with school, work, and other life experiences, the appropriate assessment task becomes that of measuring such integrated and generalized learning outcomes as the ability to cope with new problems, to find appropriate solutions, and to take the correct actions.

Measures which reflect the interdependent nature of cognitive skills essential for satisfactory functioning outside of academia have only begun to be developed.[*] For example, Klemp's General Integrative Model of Assessment incorporating a variety of independent techniques, is an approach to summative evaluation of an individual's ability to solve a problem which has as many elements and complexities of real life situations as possible. Such an assessment of individuals has the potential of coming closer to tapping real life competence than can any single test alone.

While it makes sense to require minimal levels of proficiency for many competencies, ability levels over and above necessary cut-off points do not always correlate with overall performance.

---

[*] A recent example in the noncognitive area by McClelland and Burnham (1976) reports the importance of the interaction between levels of motivation and ego-maturity for managerial competence.

For example, in a job analysis, McClelland and Dailey (1974) found that a minimal level of organizational or clerical competency was necessary for human service workers in the Massachusetts Civil Service system, but high scores on these measures were negatively correlated with superior job performance. Selecting people by rank according to score not only discriminated against those whose scores were adequate (sufficient) though "Uncompetitive," but the process failed to select the better job performers as well. This finding and others[5] suggest that going beyond sufficient levels of competency in awarding credentials can be very dysfunctional for society--not only in terms of equity, but in terms of meritocracy as well.

In many job situations, where cognitive and other competency measures are used to select job applicants, even if job relevance of the characteristics being tested for can be demonstrated (e.g., "verbal ability" in human service workers), level of sufficiency for competent job performance is rarely evaluated or known.

We need more empirical research to establish minimal levels of competence required for quality performance based on how workers in the field perform on various competency measures.

2. The Problem of Interactions

Researchers have long recognized that the interaction effects of variables are quite often more significant and meaningful than individual variables taken alone. It was stressed earlier that competence is not a simple summation of discretely defined skills and abilities. This is readily seen in the example of driving ability. Although one can identify many skills necessary for safe and effective driving--including attitudes, cognitive skills, and emotional factors, as well as perceptual and motor skills--it is intuitively obvious that a simple summation of measurement scores on these discrete task performances would not add up to equivalent driving skills. An individual who is overly competent at some driving skills but woefully inadequate in others would be a poorer driver than someone whose skills were all sufficient, though their summed skill scores would be identical.

---

[5] A recent study at Harvard revealed that the past SAT scores of faculty members were negatively correlated with more successful teachers. (McClelland, personal communication.)

The implication for higher education is that one cannot assume that abilities or skills discretely learned will be integrated in work and life functions and consequently that establishment of minimal levels of performance on isolated skills or "subcompetencies" have much meaning in themselves. Therefore, competency research, new assessment procedures, and test instruments must also focus on the interdependence of skills.

Implications of New Measures for Policy Research and Decisionmaking

The availability of measures of generic abilities which are validly linked to significant occupational and life outcomes should have important impact on educational policy, policy research, and decisionmaking. The mere establishment of the fact that abilities, which are known to be vital in the adult world [illegible] work, can now be conceptualized and measured should affect the atmosphere in which educational policy is formulated and debated. Higher education, in effect, will be put on notice that inasmuch as techniques for scrutiny are available, the processes and products of education will be scrutinized and questioned with new vigor and urgency. Time-worn answers such as "We've always done it that way," "We're building overall character and not just teaching answers to a test," or "We have no reason to believe that our program isn't working as well as any" will no longer be available to the educational administrator. Progress toward sure and solid measurability of performance may act as an improvement to that performance.

The proposed project could have two specific effects on educational policy research and decisionmaking.

1. Improvement of Means/Ends Linkages Could Be Facilitated

The availability of validated measures for assessing generic and meaningful abilities affected by education would mean that the effects of any particular program or practice could be evaluated in the same context, with increased precision and rigor. Educational policy committees continually talk about "systematic evaluation" of programs, departments, and curricular innovations. All too often such evaluations in fact turn out to involve unsystematic collection of subjective impressions at best, and *pro forma* ratifications of prejudice at worst. Many of the innovative new programs of the 1960s (as well as the abolition of traditional programs and requirements) had built-in "evaluations" after a period of a few years. In fact, however, such new programs or requirements are almost never monitored in a careful and convincing way. The availability of new measures should make it possible to evaluate existing programs and to monitor new programs with increased precision, objectivity, and thoroughness. Through careful combination of cross-sectional and longitudinal designs (see Campbell and Stanley, 1963) it will be in principle possible to establish the type and extent of contribution a particular program makes to the development of its client students.

At the same time, the financial situation that is faced by higher education both today and in the foreseeable future dictates not only that students be educated in demonstrably effective ways, but that this be done at the lowest possible cost and in the most efficient manner possible.

2. Improved Cost-Benefit Calculations Concerning Any Aspect or Part of the University Become Possible Existing ways of calculating the benefits of a particular feature of university

life (special programs, residential arrangements, or activities) often amount to crude measures such as "number of bodies processed," or "cost per student who goes through the program." As it becomes possible to specify and measure the kinds of effects supposedly produced by the program, it will then be possible to form a more realistic and useful estimate of what the institution gets for what the program costs. Again, by a combination of cross-sectional and longitudinal designs, it becomes possible to estimate the incremental improvement in the types of learning outcomes espoused by liberal arts colleges, and to distinguish this improvement from abilities already possessed at a high level by some students. This improvement can then be set against the cost of the program. In the context of a university budget, severely constrained by competing demands and limited resources, decisions about the nature and scope of programs can again be made with increased precision, objectivity, and thoroughness. For example: are special "honors" programs worth the often great additional cost in terms of faculty time and special equipment? Or are the putative "great effects" on their students more attributable to the fact that they recruit or draw students who already have the ability in question? The answer to such a question may be vitally important to the design of university policy and budgets. Such answers are simply not available so long as outcomes and effects are measured in terms of exam grades or subjective impressions.

Another example: is it possible to preserve the impact of a particular course or program while moving to media-assisted instruction and away from costly faculty-intensive discussion? The answer may be yes and it may be no; but some more precise answer to a question of that sort would have enormous impact on educational policy. A further example: are certain kinds of experiential learning techniques worth the cost? Is an expensive learner-centered program justified in terms of any measurable effect on student participants?

The simple truth is that discussion of almost any aspect of educational policy must be sharpened and made more meaningful through the availability of new kinds of measures. At the same time, these new measures should promote the ongoing development of systematic and rigorous policy research. This kind of institutional research, in turn, should improve the efficiency and effectiveness with which decisions are made in higher education.

TECHNOLOGIES FOR IDENTIFYING SKILLS, ABILITIES AND OTHER CHARACTERISTICS RELATED TO COMPETENCE

There are numerous techniques that are useful in identifying the information, skills and other characteristics necessary for successful performance as a manager and leader. But from the start we must differentiate these techniques according to three separate but important dimensions. These differentiations are critical to predicting who will be successful performers.

- We must differentiate techniques which identify critical dimensions of the job from those which identify critical characteristics of job performers.
- We must differentiate techniques that identify critical job or performer characteristics which are task, situation, or level-specific from those that identify critical job or performer characteristics which are broad or generalizable across jobs and situations and throughout a wide range of career performance levels.
- We must understand the environmental/organizational climate or dynamics within which jobs and performers interact.

The relationship among these dimensions is diagrammed in Figure 2.

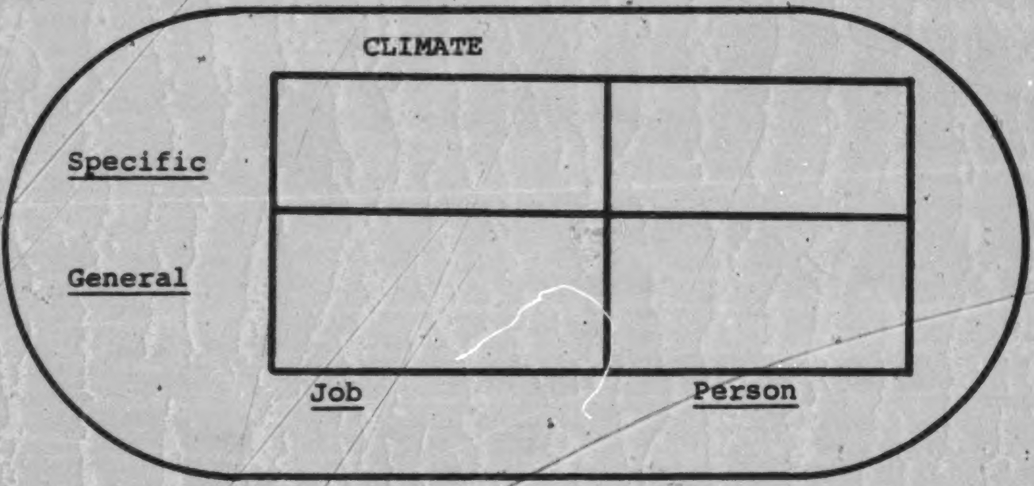


FIGURE 2

1. Job Element Analysis

The typical and/or traditional technique for identifying common or unique elements of success is to perform one or a variety of types of job function analyses. The classical approach was developed by Fine and Wiley (1971) for classifying jobs according to continuous job requirements. The job function analysis approach is based primarily on motor skills analysis and has utility in their identification but it is too narrow an approach to be used as a method for determining significant dimensions of job competence and is not related to organizational environment factors. This approach, sometimes carried to extreme, results in taxonomies of hundreds, sometimes thousands, of motor skills connected with particular kinds of jobs. These taxonomies are frequently used in developing training programs, but for other reasons besides the neglect of many significant areas of job competence, such taxonomies are not suitable guides for training. For example, there is a considerable risk of forgetting that many of these skills can be picked up on the job in a short period of time and are therefore not worthy of attention in formal career training programs. While job function analysis may help one understand common job elements for setting equitable pay scales, it does not differentiate which aspects of the job are most important to success, nor does it identify critical or differentiating characteristics of the job performer.

Flanagan and Burns (1955) moved away from the pure task-orientation approach in job function analyses by having supervisors keep a record of what they considered critical incidents involved in the work of subordinates. Whenever an employee does something that is especially noteworthy, or especially undesirable ("critical" to either good or poor performance), a notation is made in the employee's record. Over time a list of skills, abilities and characteristics that are not simply actions or action sequences is compiled. These "critical behaviors" are then classified into certain categories which can be used as rating scales. When this rating system is used supervisors note and record all "critical" instances of on-the-job behavior.

While this approach is a major revision of job function analysis, it suffers from many shortcomings. An obvious weakness is that the performance criteria identified by this method are entirely the products of subjective judgments by supervisors. Thus, criteria are severely limited by well-known perceptual screens of individual values, biases and beliefs about what should be important dimensions of the job or characteristics of job performers. Although the critical incident method offers advantages for purposes of employee counseling because it provides the supervisor with a record of behavioral observations to discuss with the employee, it does not lend itself to objective qualifications. Furthermore, there is no evidence that this approach has been used effectively for identifying managerial attributes, as opposed to those of "hourly" employees. Nor does it relate to environmental dynamics.

Primoff's (1973) Job Element Analysis is a variation of the critical incident analysis approach that bears discussion because it shows promise in filling some of the gaps left in Flanagan's clinical approach. It appears to be more systematic in its development, more quantifiable, more sophisticated in its statistical analysis and more amenable to validation. In the job element rating procedure, persons are rated on their self-reported ability to perform major elements and subelements of the job for which they are being considered.

According to Primoff, the major job elements which constitute job success include a wide variety of characteristics. Some depend on specific training; some are general. A job element may be:

- a skill, as the ability to use tools;
- an aptitude, as an aptitude for learning trade theory and practice;
- a willingness, as the willingness to do simple tasks repetitively;
- an interest, as an interest in learning new techniques;
- a personal characteristic, as reliability and dependability.

Since the purpose of the job element rating procedure is to permit evaluations of a person for the entirety of job success within a specified job classification, every aspect of job success must be included under the major elements. This is done according to three steps, as follows.

a. Tentative listing of 50-150 elements on the basis of a review of personnel rating systems.

b. Rating by experts of each tentative element in terms of relation to job success. According to Primoff, by rating the elements in terms of job success, the raters provide the same kind of information that they would if they rated people on each element and in overall success. Instead of rating people, however, they rate elements.

Elements are rated for the following four considerations:

- How important is the element for even barely acceptable work?
- How important is the element for superior accomplishment?
- How much trouble is likely if the element were to be ignored in evaluating applicants?
- How practical is it to expect applicants to be qualified in the element?

Ratings on these four dimensions are analyzed to show which five to ten elements make up success in the particular job.

c. These elements are then presented to criterion groups made up of people who fall within the job classification, one-half of whom are considered to be excellent in job performance and one-half considered satisfactory. They all rate themselves on the elements with a Self-Report Checklist.

These checklists are then numerically rated according to a Basic Crediting Plan which shows for each element the kind of evidence that would entitle the self-reporting test taker to be given a designated rating value according to the following schema."

Basic Crediting Plan for an Element	No. of Credits
Superior in an element	4
Satisfactory in the element	3
Barely acceptable (or potentially satisfactory) in the element	2
Slightly deficient in the element	1
Grossly deficient in the element	0

Primoff has developed procedures for determining the contents of each major element (termed subelements) which are used in:

- preparing an applicant checklist, rather than having him write a narrative self-report;
- amplifying the Basic Crediting Plan to fit a particular job;
- preparing a plan for a written test; and
- evaluating applicants on the checklist with the total assessment battery being used to support or contradict the items checked.

Finally, from the information about critical aspects of job performance derived from this method, in addition to the Self-Report Checklist,one can develop an aptitude test made up of elements and subelements, each with a certain weight in the test. The validity of this test is provided by a multiple regression analysis modified by Primoff and resulting in what he calls a J-coefficient. This is computed from the weights of the elements in the test and the importance of each element for a job.

There are several advantages to utilizing Primoff's procedures for identifying performance criteria over other methods described above.

- It identifies specific elements of jobs and weighs them according to their importance to job success.
- The procedure identifies aptitudes, interests and other personal characteristics not found in standard job function analyses.
- Tests can easily be constructed which tap the critical elements identified (using the J-coefficient procedure).
- The validation of critical elements is based on a comparison of superior versus average performers.
- It has a double ranking/rating procedure to increase the accuracy of ratings.
- There is a built-in flexibility for correcting errors during development.
- The self-ratings are efficient.
- Ratings can supposedly be scored reliably by one person once the Basic Crediting Plan has been completed.

While the Job Element Analysis approach has come closer to a procedure which will identify critical and quantifiable skills and abilities than other procedures discussed above, it is still reliant on expert judgment. In spite of complex and sophisticated statistical and methodological procedure for distilling these judgments into a readily usable and validated checklist, it fails to overcome the problem of eliminating perceptual screening through biased values and beliefs that may be misleading from the start.

Any judgment-based approach may indeed yield reliably observed behavioral outcomes, but may provide no insight into the skills and abilities that cause those outcomes. A clear example of this phenomenon comes from McBer's work with the U.S. Information Agency. It was universally agreed that superior U.S. Information Officers possessed a high degree of communication skills in that they were able to effectively deal with people from different nationalities and backgrounds. Communication skill per se is a criterion that could be easily rated with a high degree of reliability. However, it was found that the reason these superior officers could communicate with people so well was that they possessed two other characteristics which permitted them to do so. One was an ability; the other was an attitude. They had the ability to empathize with people, i.e., to use nonverbal cues as information and to ask questions designed to elicit the real needs of their clientele. In addition, they had a strong positive attitude toward people in general, consisting of the conviction that people are basically good and that they have the capacity to change for the better when given the means to do so. Thus, if training were only aimed toward the learning of communication skills, it would ignore the critical causal elements that are necessary for superior performance as a U.S. Information Officer. Empathy and positive bias are very difficult to measure on the job and therefore communication skills would be the desired observable criterion performance in this example. However, identifying information, skills and other characteristics necessary to achieve this criterion must often take into account attributes or characteristics which are unobservable from the point of view of a supervisor or even of the person engaged in the task of communicating.

We are not criticizing Job Element Analysis for doing what it does well, which is identifying some of the specific job requirements or personal abilities which are both observable as criteria and measurable (as predictors). However, the most appropriate use of this technique on both the criterion and predictor side of the "performance equation" relates to the analysis of very specific low level jobs or subtasks of more complex jobs.

2. Behavioral Events Analysis

McBer addresses this problem of identifying general characteristics of the person that are causally-related to complex criterion outcomes with the use of a structured interview technique. This "Behavioral Events Analysis" technique, used with success in the U.S.I.A., the Civil Service, the U.S. Navy and a variety of business and educational settings, was developed by David C. McClelland and his colleagues at McBer. It involves obtaining a number of descriptions of "behavioral episodes". For example, a senior officer might be asked to think of incidents or events in which he felt he was particularly successful, and then to describe in detail what led up to the incident, when and where it occurred, and how he was feeling and reacting before, during and after it. He would also be asked to describe incidents in which he felt he was unsuccessful or in which things did not work out the way he hoped they would. Generally, each officer interviewed would be asked to report on three successful and three unsuccessful incidents, events or episodes. Responses are recorded and analyzed by professionals experienced in this technique to "tease out" of the interview data how more effective and less effective officers perform their work differently.

A distinguishing characteristic of this interview procedure is that it elicits information from which actual behaviors can be reconstructed, rather than eliciting interpretations or perceptually biased recollections of past behavior. What further differentiates this interview approach from others is that the interviewees are initially chosen by nominations based upon job performance. The interviewees will usually fall into two categories: those who have been identified as exemplary, clearly superior, or model workers; and those who have been identified as representing an average level of competence. Differentiating incumbents into these two categories can be done in a number of ways. McBer has had much success with nominations of interviewees by supervisors who are able to view their subordinates' work under relatively standardized conditions. Although this appear to lack rigor, most supervisors asked by McBer to make nominations show a high degree of validity based upon actual behavioral and other objective performance indices. Whenever possible we include as many indices of performance which relate to measurable outcomes and peer and subordinate ratings as are available.

The advantages of the Behavioral Events Analysis are:

- It results in the identification of characteristics which are related to critical worker differences (not merely job requirement differences) and which are typically more salient or critical to high quality performance than the myriad of specific aptitudes, traits, interests, skills and other variables identified by standard job function and/or job element analysis techniques.
- It results in unique, differentiating and generalizable abilities, values and other characteristics essential to success which are otherwise perceptually screened out, as in standard interview procedures, because of naturally biased personal belief and value systems.
- It leads to specification of appropriate measures which directly underlie observable performance criteria and which are unobtainable through standard interviews, questionnaires or surveys.
- It is conceptually as well as administratively parsimonious, making it cost-effective and intuitively understandable, while gaining substantial predictive power over (or in supplement to) other techniques.

3. The Organizational Climate

Research in recent years has demonstrated that organization climate is a powerful mediator of job performance.

Campbell, Dunnette, Lawler and Weick (1970) have identified four attributes of the organizational situation: structural properties; environmental characteristics; organizational climate; and formal role characteristics. These authors defined organizational climate as:

> ...a set of attributes specific to a particular organization that may be induced from the way the organization deals with its members and its environment.

These attributes have been repeatedly shown to be closely and causally-related to leadership and work group processes and, ultimately, to factors such as satisfaction, efficiency and performance (e.g., Likert and Bowers, 1969, 1973; Franklin, 1973).

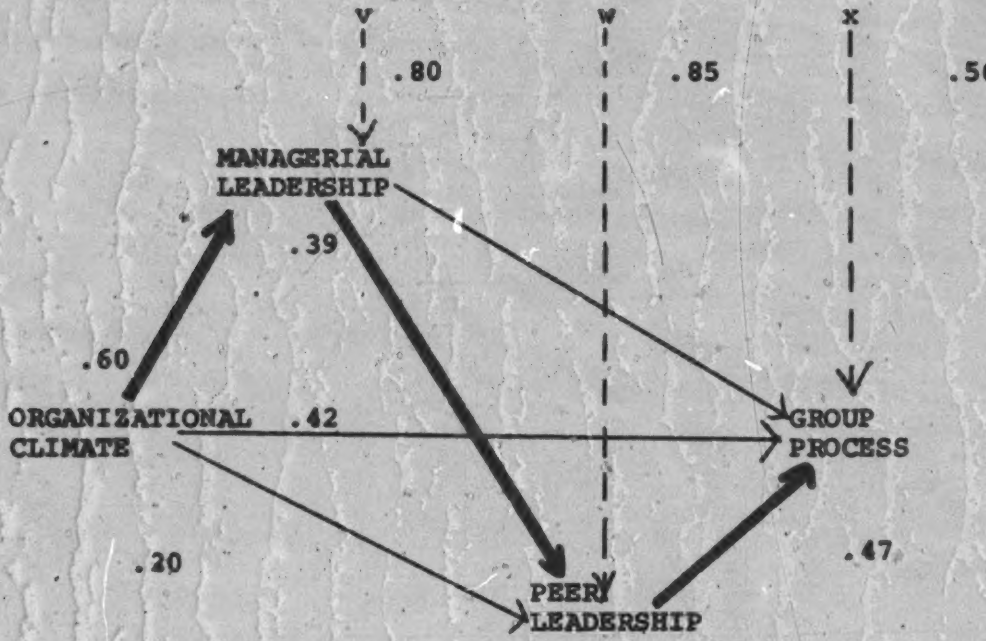
The determinants of organizational efficiency have been studied extensively in recent years, notably by Likert (1961, 1967) Likert and Bowers (1969, 1973) and Bowers and Franklin (1973). To quote Franklin (1973), "...organizational climate is the primary independent variable. Climate, along with individual differences--i.e., knowledge, skills, values--are major determinants of managerial leadership behavior which, together with organizational climate, shape peer leadership behaviors. These variables, in turn, determine group process. The final variables in this chain are individual outcomes--i.e., satisfaction, health-- and organizational outcomes--i.e., efficiency, performance, etc. (p. 19)." Implied by this discussion of the intimate link between knowledge and skill and the climate in producing effective management is the effect of new managerial skills upon the climate. As the climate is a major predictor of performance outcomes, it follows that an excellent way to assess the practical effect of a period of training on a manager is to assess the corresponding change in organizational climate.

This model was tested and verified by Bowers and Bachman (1974) who surveyed the U.S. Navy and by Franklin (1973) who drew upon a national array of civilian organizations. Their results are

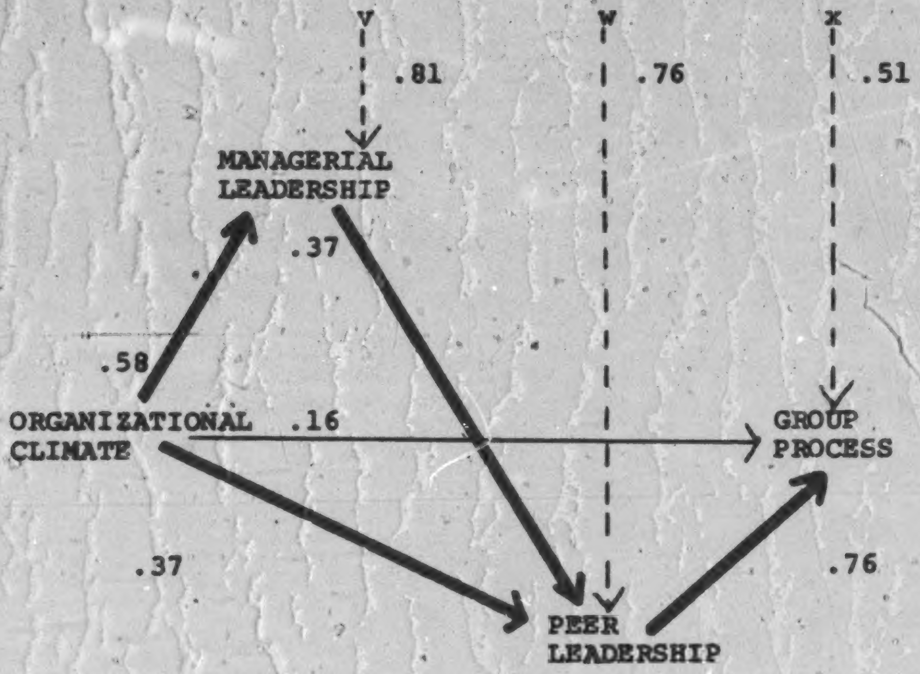
FIGURE 3

The Organizational Climate Model Fitted to Data from Civilian and Military Organizations

A. Survey of Civilian Organization


v
w
x
.80
.85
.56
MANAGERIAL LEADERSHIP
.39
.60
ORGANIZATIONAL CLIMATE
.42
GROUP PROCESS
.20
.47
PEER LEADERSHIP

B. U.S. Navy Survey


v
w
x
.81
.76
.51
MANAGERIAL LEADERSHIP
.37
.58
ORGANIZATIONAL CLIMATE
.16
GROUP PROCESS
.37
.76
PEER LEADERSHIP

━━━━ Best Predictor and Path Coefficients
———— Secondary Predictor and Path Coefficients
- - - - Residual ($1-R^2$)

attributes and job requirements on both general and specific levels in the context of overall working climate allows us to identify a comprehensive list of information, skills, values and other characteristics that lend themselves to objective measurement, differentiate superior from average performers, and provide guidelines for training and career development.

SECTION III:

PROTOTYPE MEASURES OF LEARNING OUTCOMES RELATED TO LIBERAL ARTS AND THE PROFESSIONS

We have discussed the need for new measures which (1) are sensitive and relevant to important learning outcomes of liberal arts educators, (2) have general significance to a wide variety of career and life outcomes, (3) have practical utility for educators, (4) are methodologically and technically innovative, e.g., utilizing operant rather than respondent behaviors, and (5) are quantifiable and thus amenable to rigorous determination of reliability, validity, and meaning. Using these concerns ICA has developed innovative measures which attempt to answer the need for more "proactive" (operant) measurement techniques to assess the factors of process, integration and implementation.

The purpose of this section is to present information about particular instruments which have been designed to measure competency-based outcomes. A subset of these measures is discussed in depth, and data relating those measures to academic and real world outcomes are presented. For the sake of clarity, and consistent with the competency-based orientation toward outcome-relatedness, the measures described below are organized according to three outcome domains: cognitive, effective and social outcomes.

Cognitive outcomes. Measures in this domain assess characteristics purportedly measured by traditional tests of mental ability, aptitude and knowledge. The differentiating characteristic between [illegible] measures and traditional tests is that [illegible] measures are based on the idea that the test-taker should provide all the information necessary for adequate and appropriate response to a problem on a test, as opposed to merely selecting from a set of prepared alternative responses.

Effective outcomes. Variables measured in this domain are directly translatable to behavior patterns required beyond the world of academia. This category is derived from White's (1959) term "effectance," which means positive, goal-directed and productive interaction with and influence on the environment.

Social outcomes. These measures assess areas of interpersonal competence which often facilitate the fruition of cognitive and effective dimensions of competence in life. They take into consideration the attitudes, values and orientations toward others which moderate life goals and the means for achieving them.

Discussion of Measures

Measures of Cognitive Outcomes

1. Critical Thinking. The ability to analyze new information and to synthesize new concepts based on this information reflects the ability to integrate information into one's own cognitive structure. As the cognitive structure grows, so does the ability to think critically, to make a cogent argument and to reason inductively; thus, the test of Thematic [illegible] is a

measure of cognitive development. The test takes the form of two sets of stories which an individual is asked to compare thematically. This "thematic analysis" is scored according to twelve categories of critical thinking and a total score is derived. This scoring system is reliable, efficient and cost-effective. Each scoring category is a logical and independent dimension of critical thinking skill.

This test, developed by Winter (1973), is distinguished from other measures of critical thinking skills in that it demands the test-taker to actually produce critical arguments, rather than to simply recognize the critical elements of arguments presented to him. This instrument can be used to chart a student's progress in learning this skill. Alternative versions of the test have been developed to assess both the quality and structure of critical thinking.

Recent studies undertaken to assess the effects of the college experience upon undergraduates at Wesleyan and Harvard Universities (McClelland, 1976) show that seniors score higher than freshmen on this measure. It is important to note in this context that many so-called "cognitive" tests do not reflect the improvement in students' skill over the course of a four-year college experience. When one examines firsthand the responses to the test of Thematic Analysis, however, it is not only clear that critical thinking skills improve with college, but that the scoring system for this test is intuitively satisfying in the ground it covers.

Under an ICA contract with The Fund for the Improvement of Postsecondary Education, Alverno College began to administer the test of Thematic Analysis to incoming freshmen along with other measures, including the Watson-Glaser test of critical thinking. A chief difference between Winter's measure and the Watson-Glaser is that the latter instrument only requires students to *recognize* critical thinking (a respondent measure), while the test of Thematic Analysis requires students to demonstrate critical thinking ability (an operant measure). An analysis of the data showed that the Watson-Glaser and Winter's measure of critical thinking were somewhat correlated, but only the test of Thematic Analysis was uncorrelated with respondent measures of other unrelated abilities. Those results speak favorably for Winter's measure as an uncontaminated test of critical thinking skill.

2. *Learning Styles*. A successful worker is distinguished not so much by an single set of knowledge or skills, but by the ability to adapt to and master the changing demands of one's job and career: that is, his ability to *learn*. Continuing success in a changing world requires an ability to explore new opportunities and learn from past successes and failures. Kolb's Learning Styles Inventory (1971) is a measure of individual learning styles which affect decisionmaking and problem solving. The four styles, *Concrete Experiential* learning (CE), *Reflective Observation* learning (RO), *Abstract Conceptualization* learning (AC), and *Active Experimentation* learning (AE), when present in equal proportions, indicate the type of person who

is able to involve himself fully, openly, and without bias in a new experience (CE), can reflect on and observe these experiences from many perspectives (RO), is able to create concepts that integrate his observations into logically sound "theories" (AC) and can use these theories to make decisions and solve problems (AE) (Kolb, 1973).

Extensive data has been collected on this measure in both college and postacademic settings (particularly the world of business). Kolb and Goldman (1973) have documented the utility of the Learning Styles Inventory for predicting major areas of undergraduate specialization and graduate school plans among M.I.T. undergraduates. The better the match between a student's learning style and the major subject area of the student's choice, the greater the tendency for students to place high importance in pursuing a career in that area, to perceive their workload as light, and to involve themselves with important peer groups, and the lesser the tendency for students to experience disaffection with their social and academic experience.

More ~~recent~~ work involving the analysis of administrative and technical support positions in the Division of Civil Service, Commonwealth of Massachusetts, identified "the ability to learn from experience" as a key to worker success. The Concrete Experience (CE) scale of the Learning Styles Inventory was found, in fact, to be significantly correlated with superior performance in this category of work, involving over 15 job titles (Klemp, 1976).

3. Programmed Cases. Based on incidents called from in-depth interviews with criterion groups, programmed cases can be developed to test for social learning and judgment. Versions of this technique, developed for the U.S. Information Agency and the U.S. Navy, consist of a series of incidents to which several alternative responses are attached. All of the incidents pertain to a particular individual, or "case." "Distractors," or the incorrect responses, are developed with the aid of expert judges. The cases are programmed in such a way that a person with good judgment, i.e., who does not make snap, impulsive judgments, will become more accurate in his choices of the correct alternative as he proceeds through the case.

The programmed case technology has two primary uses:

- diagnostic assessment of how one uses information in making decisions about others or predicting their behaviors, and
- examination of the process by which decisions/predictions are made, including the analysis of values, biases and preconceptions that interfere with veridical impressions of others and their situations.

These programmed cases are currently being used in psychological studies at Harvard as a measure of interpersonal learning. McBer's research interest in this technology has led to applications of programmed cases in the study of prejudice.

Klemp (1975) found that people who were exposed to cases about people whose race was unlike that of the reader were less able to predict the behavior of the person in the case than readers who were exposed to same race cases. Similar studies are planned to address the prejudicial effects of socioeconomic status and sex differences on interpersonal learning.

The direct application of programmed cases, other than personnel selection, has been in assessing the skills of human relations experts in the U.S. Navy. In a pilot study (unpublished) involving selected human resource training personnel whose performance level was known, a highly significant relationship obtained between the ability to accurately predict behavior in others, as measured by the programmed cases, and performance as a trainer in human resource management.

Other measures of cognitive outcomes, in prototype form, are the following:

4. Analysis of Argument. A test of the ability to argue for and against a controversial issue, and scored for the logical presentation of argument. (Stewart, 1976)

5. Concept Formation. A test of the ability to identify and organize similarities and differences among objects into concepts.

6. Speed of Learning. A test of how quickly one can learn new material selectively--that is, to remember functionally important information.

7. Savings Score. A test of the ability to learn new material in a particular content area--to "save" new information in an area in which the student is already well versed.

8. Proactive Case Response. A test of diagnosis, judgment, and problem solving that involves response to a detailed situation, or "case."

Measures of Effective Outcomes

1. Diagnostic Listening. The Diagnostic Listening Test consists of a taped presentation, with slides, of interviews with various individuals typical of the people one might encounter in social service work. People who take this test listen to an interview or a brief statement by a particular individual on the tape, and are then asked some questions about what has happened, what the person is really like, and what they would recommend for the person. This test requires listening, observing and judging skills which have been found necessary in human service work.

There are two subscales in this test. The Casework Subscale, consisting of 42 items, is made up of four interviews and after each of them the person taking the test is asked to answer questions and to make judgments on a multiple-choice answer sheet. The Positive Bias Subscale, consisting of 39 items, shows to test-takers three slides of clients of different sex and race with accompanying brief monologue. After each of these presentations, the test-takers are required to rate several adjectives as "does describe" or "does not describe" the client. An overall Positive Bias score is obtained by summing the number of positive yet realistic adjectives chosen. The Diagnostic Listening test measures faith in the client's ability to change, ability to observe and diagnose human problems, ability to set realistic goals, and ability to propose imaginative solutions.

Studies of human service workers in the State of Massachusetts have verified the usefulness of the skills tapped by the Diagnostic Listening Test in identifying better workers. The format of the test instrument is similar to interview situations in which workers are involved on a day-to-day basis. Both of the two subscales correlate with effective on-the-job performance as rated by supervisory consensus (McClelland and Klemp, 1974).

Introduction to Measures 2 and 3: Much research has been accumulated by McClelland (1958, 1961), McClelland and Winter (1971), Atkinson (1958), and others that shows that thought patterns are related to important kinds of behaviors. The Exercise of Imagination is McBer's version of the Thematic Apperception Test (TAT) which is used to elicit thought patterns of the test-taker.

An individual taking the test is asked to write narratives to pictures. Each of these narratives addresses the following questions about the pictures: What is happening? Who are the people? What has happened in the past that has led to the situation? What is being thought? What is wanted by whom? What will happen? and What will be done? The stories are then scored, according to a prescribed set of codes or rules, to uncover certain patterns of thought that are expressed in the stories. These scoring codes can be applied to any written narrative which addresses the types of questions mentioned above.

The link between thoughts and behavior has been repeatedly demonstrated to be strong, as opposed to the link between attitudes and behavior. The attitude-behavior link is influenced primarily by situational factors. An attitude may represent a specific goal or objective, but such goals and objectives may change according to situational demands and constraints. However, whether a specific goal changes or not, the characteristic style with which any goal is attained is determined to a large extent by thought patterns which are relatively consistent within individuals.

2. Achievement Motivation. McClelland has shown in extensive research (1961) that people high in the need for achievement are practical and interested in efficiency--in short, they are good practical decisionmakers. They are independent, good at evaluating information for its practical utility, and original in the sense that they keep looking for better ways of doing things. For instance, they make good career decisions and regularly achieve greater success earlier in their careers. In a recent Harvard University longitudinal followup study, freshmen n Ach (need for achievement) scores correlated with "early success" in various fields 14 years later (McClelland, 1976).

In the world of business, studies have shown that achievement motivation is highly related to small business success, success in sales, and performance in the role of entrepreneur (McClelland and Burnham, 1976). The need for achievement, the desire to do things better than anyone else, is particularly great among scientists and others who work against a self-imposed standard of excellence. People low in achievement motivation generally do not exhibit planning or goal-setting behavior, nor do they weigh the risks they take against expected gain. The habits of behavior in such persons may not be advantageous to success in school or in many kinds of careers. But McClelland (1965) has pointed out that people can be taught to behave in ways that are reflected by the achievement motive, and so the gap between successful performance in certain academic and work settings may be effectively bridged.

3. Self-Definition/Cognitive Initiative. Self-definition/cognitive initiative is a general characteristic of an individual which encompasses the way one thinks about the world and himself, the way one reacts to new information, and the way one behaves. People with this competency are not only able to think clearly, but also to reason from the problem at hand to a solution, and to propose and take effective action on their own. Such competence is characteristic of people who think in a rational, systematic way on their own, and who can anticipate problems before they arise. In short, it might be said that people who are high in this characteristic are able on their own to see things clearly, to understand the causes of events, to reason from problem to solution, and to take effective action to solve problems. For example, the

self-definition score has been quite useful in distinguishing between women who pursue careers following college and those who do not (Stewart and Winter, 1974).

A longitudinal study involving freshmen women at Alverno College begun by McBer with FIPSE funds, will track Self-Definition/Cognitive Initiative during the four-year college experience. The preliminary data on this measure show that it is uncorrelated with other measures of college-entry knowledge, skills, and abilities. It is therefore considered to measure a unique dimension that, because of its known predictive validity regarding the success of women in careers, is a particularly important measure in a competency-based assessment system.

Other measures of effective outcomes, in prototype form, are the following:

4. Socialized Power. A measure of whether a person is motivated to express or increase his own power for the good of the self or for the good of others.

5. Stage IV Power. A recently identified measure (McClelland, 1975) of a concern for doing one's duty, that is, to be an instrument of a power which extends beyond the self.

Measures of Social Outcomes

1. Nonverbal Sensitivity. This test, developed by Rosenthal and his associates at Harvard University (1974), consists of 40 brief voice segments on tape, all of which have been altered to obscure the words. There are two sub-

scales to the test: the RS Subscale, made up of voice segments that are randomly spliced and reassembled, and the CF Subscale, made up of segments which have been electronically filtered so that the words are unintelligible, but the intonation patterns remain. A sample item would consist of a speech segment followed by a question; e.g., "Does the segment represent someone helping a customer or criticizing someone else for being late?" Rosenthal has documented some promising criterion validity for the PONS test. High scorers on this test exhibit the following characteristics:

- they represent warmer, more honest and more satisfying peer relationships;
- they have been rated by peers and/or by teachers who know them well as being generally more sensitive in interpersonal situations; and
- they were found to be functioning more effectively in the social and intellectual areas of the California Personality Inventory.

This test, which requires less than 10 minutes to administer, has been found to predict successful performance in administrative and human service jobs, which require that the worker have "empathy," or the "ability to read between the lines" in the performance of the job (Klemp, 1976). Navy personnel involved in race relations work also have been found to score higher than the general population on this test, and the personnel who are more successful on the job also score higher than their less successful counterparts.

2. Moral Reasoning. This test is based on the research in moral development by Lawrence Kohlberg at Harvard (1970).

The test consists of a series of paragraphs which describe complex situations in which the actors are forced to choose among several moral courses of action. The task of the applicant is to write a paragraph to justify the alternative that the applicant feels is the best one on moral grounds. The essay answers are scored according to a thematic analysis developed by Kohlberg, and are interpreted according to a schema containing six levels of moral development:

Stage 1: Orientation to obedience and punishment--deference to a superior power or to trouble-avoidance.

Stage 2: Orientation to action that is satisfying to the needs of the self.

Stage 3: Orientation toward approval and to pleasing and helping others.

Stage 4: Authority and social order maintenance orientation--"doing duty" and showing respect for authority.

Stage 5: Orientation to duty defined in terms of a contract, general avoidance of violation of the rights of others, and majority will and welfare.

Stage 6: Orientation to high principle or conscience.

The conceptual categories on which the test is based have a high degree of validity as constructs.

Some recent work in the medical profession has related Kohlberg's work to the practice of physicians. High relationships exist between a physician's level of moral development and whether he will withhold or pursue treatment, the degree to which he considers the patient in the context of his family, and overall ratings of physician performance. These results show the Moral Reasoning Test to be predictive of important kinds of behavior in work which requires a good deal of value judgment. As the study of one's own values is becoming a part of what many competency-based programs wish to offer their students, Kohlberg's stage orientation to moral development is offered as an important component to this educational experience.

Other measures of social outcomes, in prototype form, are the following:

3. Affiliation Motivation. Affiliation motivation is indicated by a desire for mutual friendship; concerns with establishing, restoring or maintaining close relationships with others; and the desire to participate in friendly, convivial activities. It is an important factor in work requiring interpersonal skill and in getting people to work together as a team.

4. Social-Emotional Maturity. Abigail Stewart's measure of ego development or social-emotional maturity is based on Erickson's stage model of human behavior. Questionnaires designed to measure activities, feelings and attitudes that characterize various stages of maturity have typically had low validity, since respondent-type measures are poor indicators of behavior. By contrast, Stewart obtained the present measure of ego development by developing a coding system for the imaginative thought of individuals whose behavior placed them strongly in one of Erikson's four stages. This empirical approach conversely permits the direct classification of individuals by levels of maturity through an analysis of their written responses to the Exercise of Imagination

or similar imaginative verbal productions.

Stewart's method of classifying people into stages of ego development is based on personal physical behaviors that are easily reported and verifiable, rather than attitudes, beliefs, or preferences which are subject to bias in reporting. An additional virtue of this system is that it is the relation of behaviors to Erikson's stages, rather than a set of particular key behaviors that is important in scoring for levels of maturity. The coding system is objective and lends itself to high inter-rater reliability.

A General Integrative Model

Of the tests and measures outlined in the preceding section, none is especially useful as a diagnostic or assessment tool outside of a systematic approach to understanding the integration of the many skills that are required for success in life and work. The measures may be important pieces to the puzzle, but one cannot tell from pieces alone what the whole individual will look like. From the standpoint of competency-based education, it is the meaningful integration of life skills that is important as an outcome of the educational experience. The General Integrative Model is one way of expressing this value by involving several different measures in a system that can be used to assess student competence.

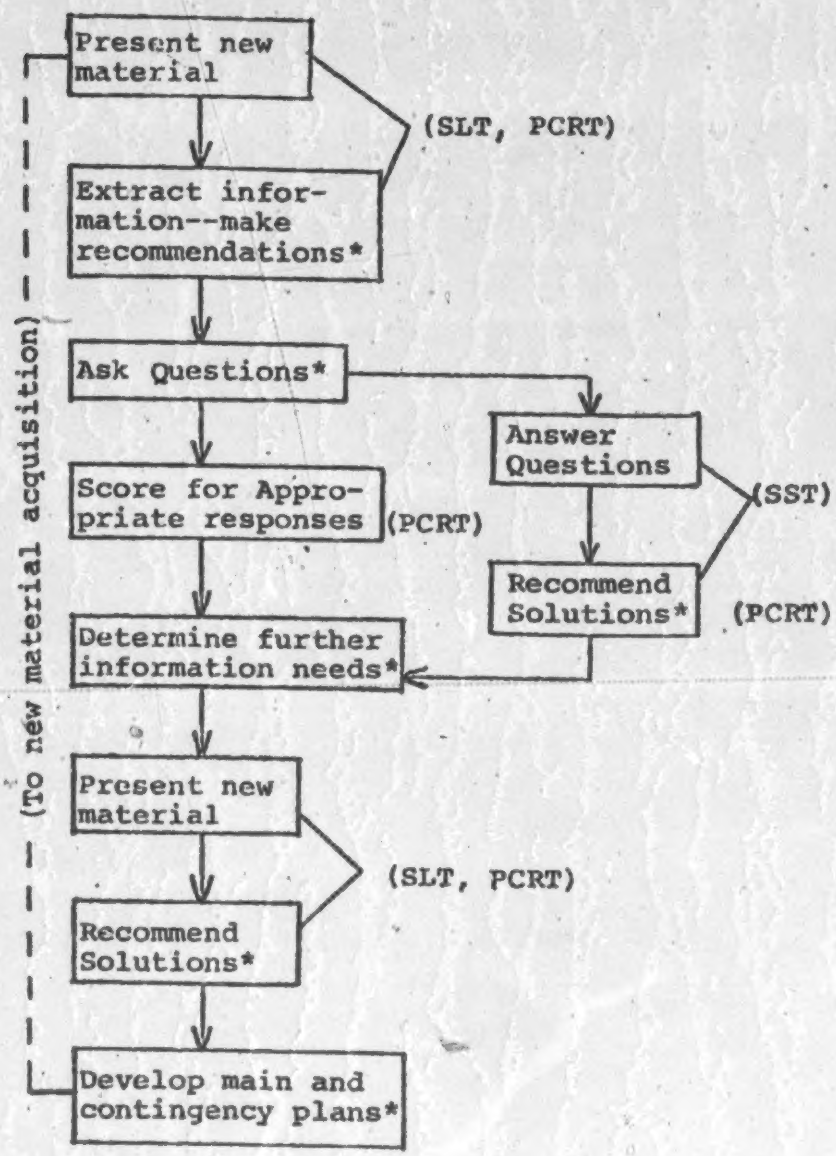
Table 1: Competency Based Measures and Their Developmental Status

	Prototype	Validated	Data Available in Academic Setting	Data Available in Real World Setting
Cognitive				
1. Critical Thinking		X	X	
2. Learning Styles		X	X	X
3. Programmed Cases		X		X
4. Analysis of Argument		X	X	
5. Concept Formation		X		
6. Speed of Learning	X		X	
7. Savings Score	X			
8. Proactive Case Response	X			
Effective				
1. Diagnostic Listening		X	X	X
2. Achievement Motivation		X	X	X
3. Self-Definition/ Cognitive Initiative		X	X	X
4. Socialized Power		X		X
5. Stage IV Power	X			
Social				
1. Nonverbal Sensitivity		X	X	X
2. Moral Reasoning		X	X	X
3. Affiliation Motivation		X		X
4. Social-Emotional Maturity		X	X	

Such general competencies as the ability to cope with new problems, to find appropriate solutions, and to take the correct action steps can be considered in such a model. Table 2 outlines one approximation to a systems approach that involves an integrated set of measures in a particular problem area, allows assessment at various junctures in the system for diagnostic purposes, and that also serves as a model for learning new skills through feedback in one's own performance. This particular version of the General Integrative Model requires an individual to demonstrate the following abilities:

- to observe;
- to extract relevant information;
- to analyze and integrate this information;
- to ask appropriate questions;
- to process new information in response to such questions;
- to utilize this information and one's knowledge in making sound and logical recommendations;
- to develop main and contingency plans;
- to set meaningful goals; and
- to feed back this new information into the process for better problem analysis and solutions.

TABLE 2: A GENERAL INTEGRATIVE MODEL
(One approximation)


Present new material
(SLT, PCRT)
Extract information--make recommendations*
Ask Questions*
Answer Questions
(SST)
Score for Appropriate responses
(PCRT)
Recommend Solutions*
(PCRT)
Determine further information needs*
Present new material
(SLT, PCRT)
Recommend Solutions*
Develop main and contingency plans*
(To new material acquisition)

SLT = Speed of Learning Test

PCRT = Pro-active Case Response Test

SST = Savings Score Test

Notes: (1) Applicable Tests are noted in parentheses at or between stages of the model.
(2) * Designates responses by the person being evaluated.

This model is not a measure _per se_, but a collection of measures logically ordered, to assess problem solving skills. The progress from stage to stage in the model presents the students with subproblems to solve, e.g., what new information to seek, what conclusions to draw, and what decisions to make derived from the information gathered at a given time.

This particular model emphasizes _cognitive_ skills, but other models can be developed that deal in different areas of competence. For example, the U.S. Navy, in their Human Goals Program, is striving to implement a training model, that uses as input tests of achievement, affiliation, and power, programmed cases, learning styles, and sensitivity to nonverbal communication. By using this model, the Navy seeks not only to assess and diagnose, but to develop curriculum aimed at more effective preparation of their personnel for work.

Characteristics and Advantages of Competency-Based Measures

This section pertains particularly to the measures outlined above, but may also be considered to be the hallmark attributes of competency-based measurement in general.

1. These tests require the person being tested to be proactive, not just reactive (i.e., one has to generate responses which can be scored for their appropriateness to real life situations). Thus, the test-taker goes beyond recognizing answers out of context. In the general model, if timing of questions or recommendations is a critical aspect of problem-

solving, then this time variable can be programmed into the model as well.

2. The tests are efficient since they can be given to groups as well as to individuals. Their efficiency and economy should substantially reduce the operational costs of current assessment procedures which require vast amounts of time, people and other resources.

3. These instruments foster equity in the assessment process, since they can be objectively and reliably scored according to the empirically validated coding systems. This is an important advantage since current methods of using juries, panels, or other groups to evaluate are not only inefficient and uneconomical, but are also vulnerable to all the vagaries of subjectivism.

4. The scores can be standardized with reference to criterion groups of which a student is preparing to become a part.

5. Many of these tests tap the competency of "learning how to learn" in a content area. This is one of the most important competencies people can develop because throughout their lives they will be faced with the problem of learning new things in selected areas.

6. These tests are much less threatening and anxiety-producing than traditional tests of recall or recognition, which because of their properties, only contribute to the fear of failure so prominent in nontraditional students.

7. A number of variations of these tests and the General

Model can be developed to add flexibility for administrators, e.g., they lend themselves to video taping, written or oral answers, individual or group testing, etc.

8. The majority of these tests have face validity. Educators and students recognize that the skills and abilities being demonstrated are applicable to general life skills.

9. Empirical and construct validation with various occupational and life skills outside of academia means that the competencies required for successful performance beyond the academic program can be established as the target of the learning process.

10. The models and tests can be validated with a variety of nonoccupation-specific populations. Some tests and models developed are noncontent-specific such that a competent person with little formal education can demonstrate competence as an analytic thinker, information processor, and a proactive initiator of appropriate solutions. The test format is easily followed and is attractive to those who are test-anxious in traditional test settings.

11. These measures can serve as pedagogical devices as well as assessment instruments, since practice in dealing with the information and component competencies necessary to solve the test problems is a direct way of learning. The instructor and student alike can easily locate and analyze weaknesses and strengths of an individual in exercising component skills. Thus, these measures can serve as diagnostic and guidance tools for supplementary curricular modules.

12. One need not take a particular course or go to a particular college in order to attain competence in the generic skills and abilities measured by these assessment tools.

REFERENCES

Atkinson, J. W. Motives in fantasy, action, and society. New York: Van Nostrand, 1958.

Bachman, J. G. and Bowers, D. G. Military manpower and modern values. Ann Arbor: Institute for Social Research, 1974.

Campbell, D. T. and Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological bulletin, 1959, 56, 81-105.

Campbell, D. T. and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Campbell, J., Dunnette, M., Lawler, E. and Weick, K. Managerial behavior, performance, and effectiveness. New York: McGraw-Hill, 1960.

Fine, S. A. and Wiley, W. An introduction to functional job analysis. Kalamazoo, Mich.: Upjohn Institute, 1971.

Flanagan, J. and Burns, R. The employee performance record: a new appraisal and development tool. Harvard business review, 1957 (September-October), 95-102.

Franklin, J. L. A path analytic approach to describing causal relationship among social-psychological variables in multilevel organizations. Ann Arbor, Mich.: Institute for Social Research, 1973.

Ghiselli, E. The validity of occupational aptitude tests. New York: John Wiley and Sons, 1965.

Hodgkinson, H. L. Evaluation to improve performance. In Vermile, D. W. (Ed.), Learner-centered reform. Current issues in higher education 1975. San Francisco: Jossey-Bass, 1975.

Klemp, G. Validation of a broad-based examination of entry-level support positions in administrative and technical work. Boston: McBer & Co., 1976.

Klemp, G. The objective measurement of prejudice and discrimination: final report to the Office of Naval Research. Boston: McBer & Co., 1975.

Kolb, D. A. On management and the learning process. Working paper #652, M.I.T., 1973.

Kolb, D. A. and Goldman, M. Toward a typology of learning styles and learning environments. *Working paper #688-73*, M.I.T., 1973.

Kolb, D. A., Rubin, I. M. and McIntyre, J. A. *Organizational psychology: an experiential approach*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Kohlberg, L. Stages of moral development as a basis for moral education. In Beck, C. and Sullivan, E. (Eds.), *Moral education*. Toronto: University of Toronto Press, 1970.

Likert, R. and Bowers, D. G. Improving the accuracy of P/L reports by estimating the change in dollar value of the human organization. *Michigan business review*, 1973, *25*(2), 15-24.

Likert, R., and Bowers, D. G. Organizational theory and human resource accounting. *American psychologist*, 1969, *24*, 585-592.

McClelland, D. C. (Ed.) *The impact of college on personal characteristics and life outcomes*. Cambridge, Mass.: Harvard University, 1976 (in press).

McClelland, D. C. *Power: the inner experience*. New York: Irvington Press, 1975.

McClelland, D. C. Testing for competence rather than for "Intelligence." *American psychologist*, 1973, *28*.

McClelland, D. C. Achievement motivation can be developed. *Harvard business review*, 1965 (November-December).

McClelland, D. C. *The achieving society*. New York: Van Nostrand, 1961.

McClelland, D. C., Baldwin, A. L., Bronfenbrenner, U. and Strodtbeck, F. L. *Talent and society*. New York: Van Nostrand, 1958.

McClelland, D. C. and Burnham, D. H. Power is the great motivator. *Harvard business review*, 1976 (March-April).

McClelland, D. C. and Dailey, C. A. *Professional competencies of human service workers: a report to the College for Human Services*. Boston: McBer & Co., 1974.

McClelland, D. C. and Klemp, G. *Report to the Executive Office of Manpower Affairs on the validation of a human service worker test*. Boston: McBer & Co., 1974.

McClelland, D. C. and Winter, D. G. Motivating economic achievement. New York: The Free Press, 1971.

Messick, S. The standard problem: meaning and values in measurement and evaluation. American psychologist, 1975, 30, 955-966.

Pottinger, P. and Klemp, G. Final report to FIPSE, Contract # OEC-0-74-9258. Boston: McBer & Co., 1976.

Primoff, E. How to prepare and conduct job element examination. Bureau of Policies and Standards, U.S. Civil Service Commission, Washington, D. C., 1973.

Rosenthal, R., Archer, D., Koivumaki, J. H., DiMatteo, M. R. and Rogers, P. Assessing sensitivity to nonverbal communication: the PONS test. Division 8 Newsletter, American Psychological Association, January, 1974.

Stewart, A. J. Scoring reactions to rhetoric of analysis of argument. Unpublished manuscript, Harvard University, Department of Psychology and Social Relations, 1974.

Stewart, A. J. and Winter, D. G. Self-definition and social definition in women. Journal of personality, 1974, 42, 238-259.

Tagiuri, R. Research needs in executive selection--a symposium. Boston: Graduate School of Business Administration, Harvard University, 1961.

White, R. W. Motivation reconsidered: the concept of competence. Psychological review, 1959, 66, 297-333.

Wilson, J. E. and Tatge, W. A. Assessment centers--further assessment needed? Personnel journal, 1973 (March), 172-179.

Winter, D. G. Scoring the test of thematic analysis for critical thinking. Unpublished manuscript, Wesleyan University, 1973.

Wollowick, H. B. and McNamara, W. B. Relationship of the components of an assessment center to management success. Journal of applied psychology, 1969, 53.